

Papel de la variabilidad genética en las enfermedades mendelianas y multifactoriales

Julián Ramírez-Bello*

Secretaría de Salud, Hospital Juárez de México, Unidad de Investigación, Ciudad de México, México

Resumen

El primer borrador de la secuencia del genoma humano, publicado en 2001, reportó gran cantidad de variantes de un solo nucleótido (SNP, single nucleotide polymorphisms). Debido a que estos polimorfismos podrían representar prácticamente toda la variabilidad involucrada en la susceptibilidad, protección, gravedad, etcétera, de diversas enfermedades comunes, así como en la respuesta de estas a los medicamentos, se pensó que podrían ser “los biomarcadores de elección” en la medicina genómica personalizada. Con la nueva información de la secuenciación de un mayor número de genomas hemos comprendido que los SNP son solo una parte importante de los marcadores genéticos involucrados en estos rasgos. Además de los SNP, se han identificado que otras variantes como las inserciones/deleciones (INDEL) y las variantes en el número de copia (CNV), las cuales — además de los clásicos repetidos en tándem de número variable (VNTR) y repetidos cortos en tándem (STR)— originan o contribuyen al desarrollo de enfermedades. El uso de estos marcadores ha servido para identificar regiones del genoma involucradas en enfermedades mendelianas (un gen-una enfermedad) o genes directamente asociados con enfermedades multifactoriales. Esta revisión tiene como objetivo describir el papel de los STR, VNTR, SNP, CNV e INDEL en los estudios de ligamiento y asociación, y su papel en las enfermedades mendelianas y multifactoriales.

PALABRAS CLAVE: Variabilidad genética. Polimorfismo. Mutación. Enfermedad multifactorial. Enfermedad mendeliana.

Abstract

The first draft of the human genome sequencing published in 2001 reported a large number of single nucleotide polymorphisms (SNPs). Given that these polymorphisms could practically represent all the variability involved in the susceptibility, protection, severity, among other aspects, of various common diseases, as well as in their response to medications, it was thought that they might be “the biomarkers of choice” in personalized genomic medicine. With the new information obtained from the sequencing of a larger number of genomes, we have understood that SNPs are only an important part of the genetic markers involved in these traits. In addition to SNPs, other variants have been identified, such as insertions/deletions (INDELs) and copy number variants (CNVs), which – in addition to classic variable number tandem repeats (VNTRs) and short tandem repeats (STRs) – originate or contribute to the development of diseases. The use of these markers has served to identify regions of the genome involved in Mendelian diseases (one gene-one disease) or genes directly associated with multifactorial diseases. This review has the purpose to describe the role of STRs, VNTRs, SNPs, CNVs and INDELs in linkage and association studies and their role in Mendelian and multifactorial diseases.

KEY WORDS: Genetic variability. Polymorphism. Mutation. Multifactorial disease. Mendelian disease.

Correspondencia:

*Julián Ramírez-Bello

E-mail: dr.julian.ramirez.hjm@gmail.com

Fecha de recepción: 29-01-2018

Fecha de aceptación: 06-12-2018

DOI: 10.24875/GMM.18004156

Gac Med Mex. 2019;155:499-507

Disponible en PubMed

www.gacetamedicademexico.com

Introducción

En 1950 se pensó que las mutaciones en el genoma humano eran eventos extremadamente raros. Un par de décadas después se identificó que el ARN se podía retro-transcribir a ADN. Esa información cambiaría nuestra forma de pensar: nuestro genoma no era estable sino dinámico. En las siguientes décadas diversos investigadores identificaron un genoma con gran variabilidad.¹ El descubrimiento de diversas mutaciones (alteraciones en el ADN con una frecuencia < 1 % en la población general) o polimorfismos (variantes genéticas comunes con una frecuencia \geq 1 % en la población general), tales como los repetidos en tándem de número variable (VNTR), repetidos cortos en tándem (STR), polimorfismos de un solo nucleótido (SNP), inserciones/delecciones (INDEL) y variantes en el número de copia (CNV) ha servido para identificar regiones (o genes ligados) o genes asociados con diversas enfermedades (Tabla 1).²⁻⁷

Variantes genéticas empleadas en los estudios de ligamiento y asociación

En los estudios de ligamiento se han empleado VNTR y STR para detectar regiones involucradas con enfermedades mendelianas (o monogénicas), tales como distrofia muscular de Duchenne, de Becker, fibrosis quística, anemia de células falciformes, etcétera, las cuales debido a su baja prevalencia resultan ser muy raras en las poblaciones. Estas enfermedades se caracterizan porque un solo gen alterado resulta en un fenotipo (Figura 1) dado que la mutación tiene un gran efecto biológico, sin embargo, siempre existen excepciones y, en algunos casos, aunque se presente una mutación esta no llevará a la enfermedad sino a una penetrancia incompleta, es decir, menos de 100 % de los individuos presenta el fenotipo esperado según su genotipo (Tabla 1 y Figura 1).

El ligamiento (paramétrico) se refiere a la cosegregación de estos marcadores con una enfermedad o fenotipo en familias con múltiples miembros afectados; se mide en logaritmo de base 10 (*LOD score*) y debe mostrar al menos un *LOD score* de 3, lo que significa que el marcador genético muestra un ligamiento de 1000 veces con el rasgo de interés. Así, una señal positiva (*LOD score* \geq 3) de ligamiento sirve para identificar regiones cromosómicas (aunque no necesariamente el gen) involucradas en las patologías mendelianas (Figura 1).⁸⁻¹⁰ Dado que esta

estrategia no detecta al gen ligado, sino más bien una región de un cromosoma, el siguiente paso es detectar la mutación en un gen cercano implicado con la enfermedad mediante otras metodologías (Figura 1).

Por su parte, en los estudios de asociación, donde se observa la co-ocurrencia de un marcador genético con una enfermedad o fenotipo de interés,^{8,9} se emplean los VNTR, STR, SNP, CNV e INDEL con el fin de detectar genes asociados con el rasgo de interés (Figura 1). Dada su abundancia en el genoma y relativa facilidad para evaluarlos por métodos automatizados, los SNP son usados rutinariamente en este tipo de análisis.^{6,7} Su empleo en las investigaciones de casos-controles es muy conveniente (en este artículo solo se hace alusión a los estudios de casos-controles), ya que se puede hacer un mapeo genético mucho más fino *versus* los VNTR o STR, escasos si se comparan con los SNP. Algunos SNP pueden ser evaluados mediante estudios de genes candidato (genes que codifican o no proteínas y que por su efecto biológico participan en la patogénesis de una enfermedad) o miles pueden ser examinados mediante estudios de asociación del genoma completo (GWAS), los cuales detectan decenas de *loci* relacionados con esos rasgos (en estos estudios no se establecen hipótesis debido a que no están dirigidos a ningún gen en particular).

Es importante aclarar que para tener resultados confiables en los GWAS, estos deben incluir cientos o miles de muestras de casos y controles y replicarse en al menos un segundo grupo de estudio, debe haber correcciones e incluirse la prueba de Bonferroni (si se evalúa un millón de SNP, entonces se tomarán valores de $p = 0.05/10^{-6}$, es decir 5×10^{-8}), la corrección por ancestría (mediante marcadores informativos de ancestría), etcétera; también se sugiere alguna prueba funcional de la variantes asociadas.¹¹ Los estudios de ligamiento y asociación presentan diversas diferencias (Tabla 1).

Mutación o polimorfismo

En las enfermedades mendelianas o multifactoriales, las mutaciones y los polimorfismos, aunque representan un cambio en la secuencia de ADN, son definidas en forma diferente debido a su frecuencia alélica y efecto biológico en un individuo (Figura 2). En las enfermedades mendelianas se mencionan las mutaciones, mientras que en las enfermedades multifactoriales se habla de polimorfismos. En las primeras basta una sola mutación en un gen (con una

Tabla 1. Características de las enfermedades mendelianas y multifactoriales (en estas solo se mencionan rasgos de los estudios de casos-contrroles no relacionados)

Rasgos	Enfermedad mendeliana	Enfermedad multifactorial
Objetivo	Identificación del gen o genes principales ligados en familias	Identificación del gen/genes de bajo tamaño del efecto asociados con una base poblacional
Alteración genética	Mutación	Polimorfismo
Modelo de herencia	Bien definido Autosómico recesivo, autosómico dominante, ligado al X, disomía uniparental, etcétera	No definido
Efecto de la alteración	Alto Generalmente, todos los individuos que presentan la mutación en un único gen, también presentan la enfermedad (excepto en la penetrancia incompleta)	Bajo Los individuos que presentan todas las variantes asociadas no desencadenan la enfermedad, causan susceptibilidad para las enfermedades
Frecuencia de la alteración	Rara Menor a 1 % en una población	Bajas-comunes Con una frecuencia de al menos 1 % en la población general
Variantes empleadas en estos estudios	VNTR, STR, SNP	VNTR, STR, SNP, CNV, INDEL
Efecto del medio ambiente	Prácticamente nulo	Fuertemente influidas
Presentación clínica	Se presentan en etapa más temprana de la vida o durante la niñez	Se presentan frecuentemente más hacia la etapa de adultos jóvenes o mayores de edad. Otras, como la obesidad y diabetes tipo 2, actualmente se presentan también en niños

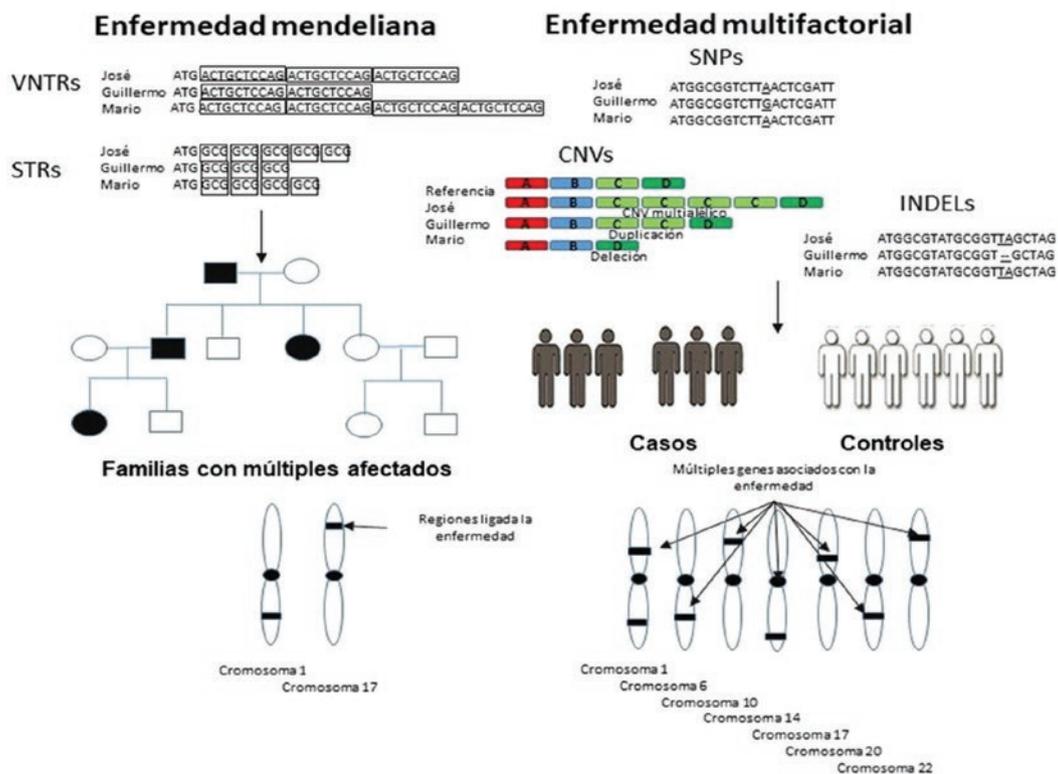


Figura 1. Identificación de regiones o genes ligados mediante VNTR, STR, SNP, CNV e INDEL en familias o genes asociados en individuos no relacionados, en diferentes enfermedades o rasgos de interés. Los VNTR y STR se han empleado para mapear regiones cromosómicas ligadas con diversas enfermedades mendelianas, donde un único o pocos genes con gran penetrancia son los responsables de los fenotipos observados. En los estudios de asociación genética, VNTR y STR también fueron utilizados para identificar genes con un tamaño del efecto pequeño, sin embargo, por su baja distribución dentro del genoma y de los genes, los marcadores más utilizados en estos estudios son los SNP, aunque también se emplea CNV e INDEL.

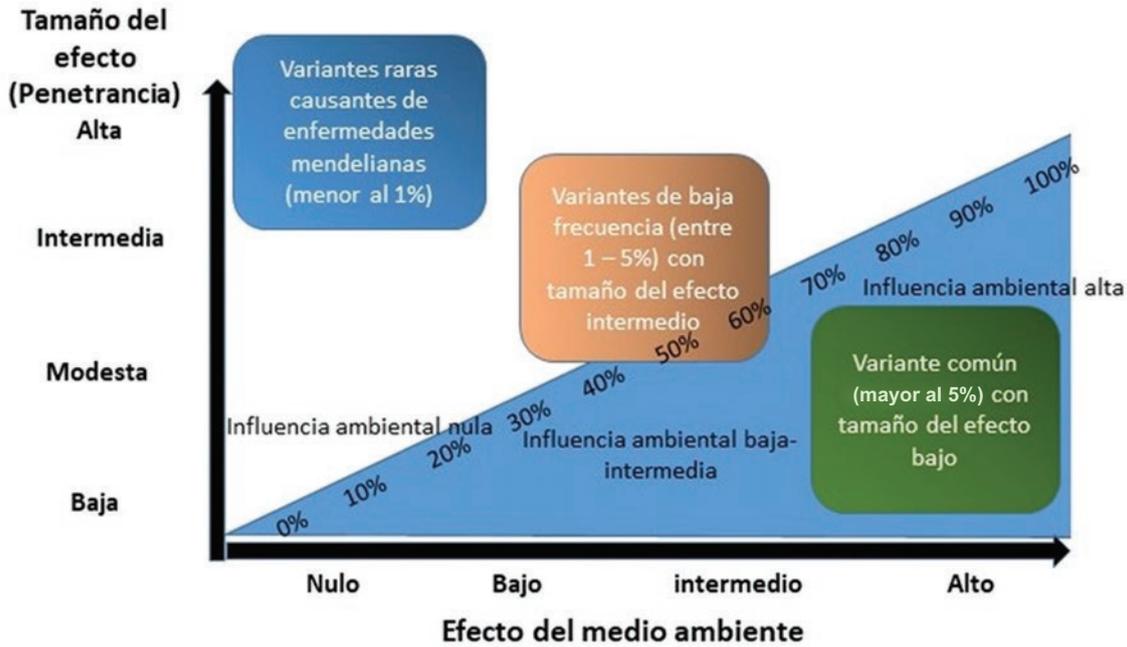


Figura 2. Imagen representativa del efecto del medio ambiente y de las mutaciones-polimorfismo en la génesis de las enfermedades mendelianas y multifactoriales. El medio ambiente tiene poco o nulo efecto en la génesis de la mayoría de las enfermedades mendelianas (causadas por mutaciones); por el contrario, puede influir hasta casi en 100 % sobre las enfermedades multifactoriales (múltiples genes y con un tamaño del efecto pequeño), mientras que las variantes génicas pueden tener poca o mucha relevancia en el desarrollo de estas enfermedades.

frecuencia < 1 %, la cual no se presenta en la población general) para que se presente fibrosis quística, distrofia muscular de Duchenne o de Becker, entre otras (excepto en la penetrancia incompleta), y prácticamente en ellas no influye el medio ambiente (Figura 2). En estas enfermedades, las mutaciones siguen un patrón de heredabilidad bien definido e incluyen a la herencia autosómica recesiva, autosómica dominante, ligada al X, etcétera.¹²

Por su parte, las enfermedades multifactoriales están influidas por decenas o centenas de genes (poligénicas) de susceptibilidad y cada *loci* contribuye con un tamaño del efecto pequeño (Figura 2). Adicionalmente, los alelos de las variantes asociadas presentan frecuencias que van de bajas a comunes (Figura 2), sin un patrón de herencia bien definido (por este rasgo se definen también como enfermedades complejas) y en su desarrollo están implicados factores de riesgo ambientales, es decir, un individuo aun con todos los alelos de riesgo no presentará la enfermedad a menos que exista un factor de riesgo ambiental que dispare su desarrollo (Tabla 1 y Figura 2),^{6,7} por ejemplo, en algunas enfermedades autoinmunes, como artritis reumatoide y lupus eritematoso sistémico (LES), se han identificado más de 100 genes asociados con susceptibilidad, sin embargo, estos no son suficientes para desencadenarlas.^{13,14}

VNTR en enfermedades mendelianas y multifactoriales

Los VNTR, conocidos también como minisatélites, comprenden repetidos en tándem de secuencias no codificantes de 10 a 100 nucleótidos de longitud (Figura 1). Se ha calculado que existen aproximadamente 1500 VNTR en el genoma humano.^{1,4,10} Su uso como marcadores para estudios de ligamiento se dio en los inicios de la década de 1980, cuando fueron empleados para mapear regiones cromosómicas ligadas con diversas enfermedades mendelianas (Figura 1). Los VNTR tuvieron utilidad en la identificación de “señales” cercanas a los genes principales causales de estas enfermedades; estudios posteriores identificaron esos genes.^{8,10}

Por otro lado, el uso de VNTR en los análisis de ligamiento para enfermedades complejas no tuvo el éxito esperado, dado que estas enfermedades están influidas por varios genes con un tamaño del efecto pequeño (no hay un gen principal).⁸ Adicionalmente, los VNTR han sido empleados en los estudios de asociación genética, aunque en la actualidad han quedado prácticamente olvidados debido principalmente a su bajo número y porque la gran mayoría se localiza en los extremos cromosómicos. Sin embargo, no estudiarlos implica potencialmente “perder” (heredabilidad perdida) algunos *loci* involucrados en

Tabla 2. Variantes genéticas involucradas en enfermedades mendelianas y multifactoriales

Variante	Enfermedad mendeliana	Enfermedad multifactorial	Referencias
VNTR	Hipertensión arterial	Diversos tipos de cáncer, artritis reumatoide, LES, osteoartritis	1, 15-18
	Regiones cromosómicas ligadas: 2p22.1, 5q33	Genes: <i>HRAS</i> , <i>IL1</i> , <i>IL4</i> , <i>DNasa 1</i>	
STR	En más de 40 enfermedades mendelianas se han ligado diversos STR; entre las enfermedades se encuentran ataxia espinocerebelosa, enfermedad de Huntington, atrofia muscular espinobulbar, distrofia miotónica	Diversos tipos de cánceres, enfermedades autoinmunes	3, 19-30
	Gen de la hungtintina, <i>ATXN3</i> , <i>DM1</i>	Genes: <i>KCNQ1OT1</i> , <i>PTPN11</i> , <i>ECRG2</i> , <i>TCR</i>	
SNP	—	Diversos tipos de enfermedades autoinmunes, metabólicas, etcétera.	6, 7, 38, 39
	—	Genes: <i>PTPN22</i> , <i>TNF</i> , <i>FCRL3</i> , <i>microRNA</i> , <i>FTO</i> , <i>LEPR</i> , etcétera	
CNV	Microtia autosómica dominante, enfermedades del neurodesarrollo como autismo y epilepsia, defectos cardíacos congénitos, etcétera	Diversos cánceres, enfermedades autoinmunes, mentales, etcétera	40-43
	Regiones cromosómicas identificadas: 4p16	Genes: <i>FCGR3B</i> , <i>C4</i> , <i>CCL3L1</i> , <i>NBPF</i> , <i>UGT2B17</i> ,	
INDEL	—	Esclerosis múltiple, LES, obesidad e índice de masa corporal	46, 48
	—	Genes: <i>BAFF</i> , <i>LEPR</i> , <i>UCP2</i>	

VNTR = repetidos en tándem de número variable, STR = repetidos cortos en tándem, SNP = polimorfismos de un solo nucleótido, CNV = variantes en el número de copia, INDEL = inserciones/delecciones, LES = lupus eritematoso sistémico, HRAS = proto-oncogen, IL1 = interleucina 1, IL4 = interleucina 4, ATXN3 = ataxina 3, KCNQ1OT1 = transcrito 1 hebra antisentido de KCNQ1, PTPN11 = fosfatasa de tirosinas no receptor tipo 1, ECRG2 = gen 2 relacionado con cáncer de esófago, TCR = receptor de células T, PTPN22 = proteína tirosina fosfatasa no receptora tipo 22, TNF = factor de necrosis tumoral, FCRL3 = proteína parecida al receptor de la fracción cristalizante 3, FTO = proteína asociada con obesidad y masa magra, LEPR = receptor de leptina, FCRL3 = RECEPTOR 3 B de la fracción cristalizante gamma, C4 = proteína 4 del complemento, CCL3L1 = proteína 1 parecida al ligando de quimiocina C-C3, NBPF23 = pseudogen de NBPF, UGT2B17 = UDP glucuronosiltransferasa de la familia 2, miembro 17; BAFF = factor activador de células B, UCP2 = proteína desacoplante 2.

este tipo de enfermedades. Se ha propuesto que un VNTR de 28 pb localizado en el oncogen *HRAS* causa susceptibilidad para varios tipos de cáncer a través de los siguientes mecanismos:

- El desequilibrio de ligamiento, es decir, algunos alelos de este VNTR están en desequilibrio de ligamiento con la verdadera variante asociada en esos tumores.
- La afectación de la expresión génica de factores de transcripción rel-NF-kB, dado que se localiza en un sitio de unión para esa familia de factores de transcripción y podría ser causante directo de la asociación.¹

Por otro lado, recientemente se identificó un VNTR con cuatro repetidos de 27 pb localizado en el exón 4 del gen sintasa de óxido nítrico endotelial (*NOS3*) que confiere susceptibilidad para enfermedad arterial coronaria. Los estudios de correlación mostraron que este VNTR está asociado con niveles altos de triglicéridos en suero (*versus* el de cinco repetidos de 27 pb), mientras que el análisis *in silico* mostró que este repetido puede afectar el corte y empalme alternativo.¹⁵

Se ha observado que otros VNTR localizados en genes de mioglobina, del antagonista del receptor de interleucina-1, DNasa 1, interleucina-4, entre otros, están asociados con artritis reumatoide, LES y osteoartritis (Tabla 2).¹⁶⁻¹⁸

STR en enfermedades mendelianas y multifactoriales

Los STR (conocidos también como microsatélites) son repetidos en tándem que oscilan entre 2 y 10 pb, aunque algunos autores los definen como repetidos de 2 pb (Figura 1). Debido a su amplia distribución en el genoma humano, alcanzan hasta 3 % del total de nucleótidos, representando cerca de 6 % de las regiones codificantes (donde se incluyen las regiones 5' no traducidas, secuencia codificante y 3' UTR).³ Recientemente se ha reportado un primer catálogo de STR humanos a escala de estudios poblaciones amplios del genoma: se han identificado cerca de 700 000 STR en 1009 individuos en la fase 1 del proyecto de los 1000 genomas, hallazgo sobresaliente debido a

que contribuirá a identificar parte de la “heredabilidad pérdida” en las enfermedades multifactoriales.³

Los STR se encuentran en diferentes partes de la estructura de los genes codificantes de proteínas y pueden tener un efecto en la expresión génica, estabilidad de los ARN y creación o destrucción de un marco de lectura abierto, etcétera.³ Presentan alta hipermutabilidad, es decir, los repetidos tienden a generar más copias debido a errores de recombinación o replicación. Estas variantes han sido usadas como biomarcadores genéticos en los estudios de ligamiento, con más éxito que los VNTR debido a su mayor distribución.⁸ Por ejemplo, algunos estudios en artritis reumatoide (una enfermedad compleja y multifactorial) mostraron que varios STR localizados en el cromosoma 6p21 son importantes para esta patología, al mismo tiempo, con estos marcadores se identificaron otras regiones ligadas en esta enfermedad autoinmune.^{19,20}

Por otro lado, diversos STR han sido ligados a más de 40 enfermedades mendelianas, incluidas la ataxia espinocerebelosa, enfermedad de Huntington, atrofia muscular espinobulbar, entre otras, las cuales se originan por una dramática expansión de trinucleótidos y se comportan como mutaciones dominantes de ganancia de función.^{3,21}

El STR CAG (codón implicado en la codificación de glutamina) localizado en la secuencia codificante del gen de huntingtina ha sido identificado como el responsable de la enfermedad de Huntington (Tabla 2); en 1993, en individuos sanos se identificaron menos de 35 repetidos, mientras, los afectados presentaron 37 o más repetidos de glutamina, lo cual lleva a un cambio en la estructura y función de la proteína.^{22,23} Otro ejemplo lo representa la ataxia espinocerebelosa tipo 3 (Tabla 2), en la que también una expansión del trinucleótido CAG en la secuencia codificante del gen *ATXN3* (codificante de la proteína ataxina 3) se asocia con un plegamiento anormal de la proteína y con una acumulación en ciertas regiones del cerebro.²⁴ Por otro lado, el repetido CTG localizado en el gen que codifica para la proteína cinasa MD1 ha sido implicado en la distrofia miotónica tipo 1 y el repetido GGGGCC del gen hipotético *C9ORF72* está involucrado en la aparición de la demencia frontotemporal.²⁵

Respecto al papel de STR en enfermedades multifactoriales, varios de ellos han sido asociados con susceptibilidad, por ejemplo, el STR localizado en el gen *KCNQ1OT1* (gen de ARN largo no codificante involucrado en el silenciamiento de un agrupamiento de genes en cis) causa disminución en el riesgo para

el desarrollo de carcinoma hepatocelular.²⁶ Otros STR localizados en genes como *PTPN11*, *ECRG2* y *TCR* han sido asociados con el desarrollo de diversos cánceres y enfermedades autoinmunes (Tabla 2).²⁷⁻³⁰

SNP y enfermedades multifactoriales

Los SNP son variantes de un solo nucleótido, generalmente bialélicos, y representan los polimorfismos más abundantes del genoma humano (se encuentran cada 200 a 500 pb). Datos recientes del Proyecto 1000 Genomas indican que hay cerca de 11 millones de ellos y cada individuo presenta en promedio tres millones.^{5,31,32} Dada su amplia distribución, los SNP se encuentran en regiones donde hay o no genes. En los primeros se localizan en cualquier parte de su estructura e incluye los promotores, exones e intrones, mientras que los SNP localizados en genes de ARN no codificantes pueden encontrarse en los promotores o en la secuencia no codificante.⁷ Desde el punto de vista funcional, los SNP localizados en los genes codificantes y no codificantes de las proteínas afectan diversos procesos que se mencionan con detalle en el artículo “Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multifactorial diseases”.⁷

Desde su descubrimiento, los SNP han sido empleados en los estudios de asociación.^{7,33} Las asociaciones entre estas variantes y los rasgos multifactoriales se miden por la significancia estadística, la cual generalmente es de $p \leq 0.05$ para estudios de gen candidato o de $p \leq 5 \times 10^{-8}$ para GWAS.^{11,34} Sin embargo, es importante tomar en cuenta ciertas deficiencias en las investigaciones de gen candidato, las cuales incluyen estratificación poblacional, bajo tamaño de muestra, bajo poder estadístico, etcétera.^{9,11,33} las cuales no representan un problema en los GWAS, cuyo análisis incluye evaluar diversos marcadores de ancestría, un tamaño de muestra alto (cientos o miles de casos-controles), réplicas de los hallazgos en otras muestras, así como la eliminación de otros sesgos que pueden confundir las asociaciones.¹¹ Mediante ambas estrategias se han reconocido múltiples genes asociados con enfermedades multifactoriales,^{6,7,11,33} por ejemplo, un estudio de gen candidato identificó que SNP en los pre-miARN *146a* y *499* confiere susceptibilidad para LES,³⁵ mientras que el SNP no sinónimo C1858T —el cual lleva a un cambio de arginina por triptófano en la posición 620 del mRNA del gen *PTPN22*— confiere susceptibilidad

para artritis reumatoide³⁶ y para enfermedad de Graves³⁷ ya que afecta la estructura y función de la proteína, así como diversos procesos inmunológicos.³⁶ El alelo C del SNP-169T/C de *FCRL3*—que codifica para la proteína 3 parecida a los receptores de la fracción cristalizable gamma— modifica la afinidad de unión para el factor de transcripción NF-κB y se correlaciona con mayores niveles de mRNA de este receptor implicado en la activación de las células B, causando susceptibilidad para LES y artritis reumatoide (Tabla 2).^{6,7}

Por otro lado, un análisis reciente de 225 SNP anotados en una base de datos de polimorfismos previamente asociados con obesidad, identificó que 15 variantes del gen *FTO* están asociadas con el índice de masa corporal, por lo que replicó hallazgos previos y confirmó que *FTO* representa el principal *locus* de riesgo para este rasgo.³⁸ Aunque estas variantes se encuentran en el intrón 1 de este gen, estudios funcionales muestran que algunos SNP pueden formar conexiones de largo alcance para regular la expresión del gen *IRX3*, el cual ha sido implicado en algunos eventos relacionados con obesidad a través de *FTO*.³⁸ Otro estudio ha mostrado el papel de diversos SNP localizados en el gen del receptor de leptina (*LEPR*) y su relación con la susceptibilidad para obesidad.³⁹ Existen otros ejemplos acerca del papel de SNP y su relación con enfermedades multifactoriales.^{6,7}

CNV en enfermedades mendelianas y multifactoriales

Las CNV (un tipo de variación estructural) abarcan segmentos de ADN de una longitud igual o mayor a 1 kb, las cuales se presentan en un número variable de copias entre individuos y comprenden deleciones, duplicaciones e inserciones (Figura 1).⁴⁰ Se ha sugerido que las CNV influyen hasta en 1 % de la variabilidad entre dos individuos, mientras, los SNP representan cerca de 0.1 %.⁴⁰ Además, las CNV pueden causar enfermedades mendelianas y susceptibilidad en las enfermedades multifactoriales.⁴⁰⁻⁴² Desafortunadamente, la cantidad de CNV identificadas en el humano no está definida, dado que depende de la naturaleza del arreglo que se use para evaluarlas. Actualmente en la base de datos de CNV hay depositadas 29 133, 41 % se sobrepone con genes conocidos, esto significa que una parte importante de ellas pueden causar enfermedades mendelianas o multifactoriales.⁴³ Desde el punto de vista efecto biológico-patológico, las CNV son las variantes más

importantes (debido a su tamaño) relacionadas con susceptibilidad para diversas enfermedades humanas (cuando se comparan con VNTR, STR y SNP). Se ha reportado que causan enfermedades mendelianas en el periodo neonatal o en etapas tempranas de la niñez, incluidas enfermedades del neurodesarrollo como autismo o epilepsia y defectos cardiacos congénitos, entre otras (Tabla 2).⁴³

Respecto a las enfermedades multifactoriales, recientemente se mostró que una CNV de 8 kb localizada en el gen *NEGR1* está asociada con protección para obesidad y un estudio funcional indicó que el factor de transcripción NKX6 se une a esta región y reprime su transcripción, por lo que la deleción de esta CNV evita la unión de este factor de transcripción. Aún se requieren análisis adicionales para comprender mejor su papel en obesidad.³⁹

Por otra parte, varias CNV localizadas en *FCGR3B*, *C4*, *CCL3L1*, y *DEBF* han sido asociadas con el desarrollo de glomerulonefritis en pacientes con LES, VIH/sida y psoriasis, respectivamente (Tabla 2).^{2,41-43} Es importante mencionar que los datos de asociación genética de CNV localizadas en *FCGR3B*, *CCL3L1*, *DEBF* y *C4* deben tomarse con mucha precaución dado que estas asociaciones deberían ser identificadas con estudios de microarreglos de SNP, sin embargo, hasta ahora no hay evidencia de asociación de estas variantes mediante estas herramientas moleculares.

INDEL y enfermedades multifactoriales

Recientemente, han sido reportadas pequeñas INDEL (Figura 1) en el genoma humano, las cuales comprenden una longitud de 1 a 10 000 pb.⁴⁴ Un estudio en el cromosoma 22 mostró que de 100 % de variantes, 13 % correspondía a INDEL.⁴⁴ Otra investigación llevado a cabo en 330 genes de diferentes individuos reportó 2393 pequeñas INDEL de 1 a 543 pb de longitud.⁴⁴ Por otro lado, se identificaron 3.4 millones de pequeñas INDEL, 1.96 millones fueron no redundantes, mientras que 819 363 se localizaron en genes humanos, de las cuales 2123 se encontraron en los exones. El número final de INDEL identificadas en el genoma tampoco no está establecido, ya que depende de la herramienta de última generación que se utilice y la cantidad de INDEL que cada persona presenta, por ejemplo, el genoma de Watson contiene 222 718, mientras, el de Venter posee 823 396.⁴⁴

Dada su reciente identificación, las INDEL no han sido evaluadas en análisis de ligamiento, pero sí en

análisis de asociación. Desde el punto de vista funcional y dado que las pequeñas INDEL (sobre todo las que abarcan algunas décimas de pb) pueden encontrarse en genes y afectar a promotores, exones (5' UTR, secuencia codificante, y 3' UTR) e intrones, es posible que influyan en el desarrollo de enfermedades multifactoriales debido a que también son capaces de alterar aspectos de los mRNA o proteínas, por ejemplo, 184 INDEL de 2123 se encuentran cercanas a los sitios de unión entre exones y intrones, afectando el corte y empalme y a mRNA-proteínas útiles para las células; las restantes 1939 se localizan en la secuencia codificante y provocan alteraciones en el corrimiento en el marco de la lectura o en la terminación prematura de la proteína.^{45,46}

Respecto al papel de estas variantes y su asociación genética con algunas enfermedades, se ha reportado que la INDEL GCTGT-> A (el alelo A es de riesgo) localizada en la región 3' UTR del gen *TNFS-F13B* (el cual codifica para la citocina factor activador de células B) confiere riesgo para desarrollar esclerosis múltiple y LES en pacientes de diversos países de Europa (Tabla 2).⁴⁷ Es importante mencionar que la variante A de la INDEL GCTGT->A genera un mRNA corto que escapa de la inhibición de un microARN (un regulador negativo de la expresión de diversos mRNA a nivel postranscripcional), lo cual lleva a mayor producción de citocina factor activador de células B soluble en estos pacientes, causando susceptibilidad para su desarrollo.⁴⁷

Por otro lado, la INDEL CTTTA localizada en la región 3' UTR del gen *LEPR* está asociada con aumento de peso corporal,⁴⁸ mientras que la INDEL de 45 pb localizada en la región 3' UTR del gen *UCP2* (que codifica para una proteína desacoplante que actúa como un transportador mitocondrial y regulador de la homeostasis energética y termogénesis) ha sido asociada con alteraciones en el índice de masa corporal y el peso corporal después de la alimentación.⁴⁸

Otro estudio evaluó un grupo de INDEL en diversos genes de pacientes con cáncer colorrectal; se detectó que estas variantes en los genes *ACE*, *UCP2*, *TYMS*, *IL4*, *NFKB1*, *CASP8*, *TP53*, *HLAG*, *UGT1A1* y *SGSM3* están asociadas con susceptibilidad y algunos rasgos clínicos (Tabla 2).⁴⁹ Otras pequeñas INDEL localizadas en los genes *PPP3R1*, *PARP1*, *pre-miR3131*, *COL1A2*, *HLA-G* han mostrado asociación con susceptibilidad para enfermedad arterial coronaria, melanoma, carcinoma hepatocelular, osteoporosis y LES, respectivamente.⁵⁰⁻⁵⁴

Conclusiones

Marcadores genéticos como VNTR, STR, SNP, CNV e INDEL han sido útiles en la identificación de genes relacionados con enfermedades mendelianas y multifactoriales. Desde su aparición, los VNTR y STR fueron empleados en estudios de ligamiento para identificar regiones involucradas con enfermedades raras, sin embargo, en las comunes no fueron exitosos. Con la identificación posterior de los SNP, estos se convirtieron en los marcadores genéticos de elección para realizar estudios de asociación genética. Actualmente, con el análisis de las INDEL y CNV mediante diversas herramientas de baja, mediana y alta densidad hemos comenzado a comprender su papel en la génesis de enfermedades mendelianas y en la susceptibilidad para enfermedades multifactoriales. En un futuro no muy lejano, conocer la variabilidad genómica involucrada en ambos tipos de enfermedades tendrá como consecuencia el desarrollo de una medicina individualizada, preventiva y predictiva.

Bibliografía

1. Ramel C. Mini-and microsatellites. *Environ Health Perspect.* 1997;105:781-789.
2. Haraksingh RR, Snyder MP. Impacts of variation in the human genome on gene regulation. *J Mol Biol.* 2013;425:3970-3977.
3. Press MO, Carlson KD, Queitsch C. The overdue promise of short tandem repeat variation for heritability. *Trends Genet.* 2014;30:504-512.
4. Richard GF, Kerrest A, Dujon B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. *Microbiol Mol Biol Rev.* 2008;72:686-727.
5. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. *Nature.* 2010;467:1061-1073.
6. Ramírez-Bello J, Vargas-Alarcón G, Tovilla-Zárate C, Fragoso JM. Single nucleotide polymorphisms (SNPs): functional implications of regulatory-SNP (rSNP) and structural RNA (srSNPs) in complex diseases. *Gac Med Mex.* 2013;149:220-228.
7. Ramírez-Bello J, Jiménez-Morales M. Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multifactorial diseases. *Gac Med Mex.* 2017;153:238-250.
8. Ramírez-Bello J, Pérez-Méndez O, Ramírez-Fuentes S, Carrillo-Sánchez S, Vargas-Alarcón G, Fragoso JM. Genetics and genomic studies in the hypertension: an actualization of the genomic studies. *Arch Cardiol Mex.* 2011;81:240-250.
9. Baron M. The search for complex disease genes: fault by linkage or fault by association? *Mol Psychiatry.* 2001;6:143-149.
10. Nakamura Y. DNA variations in human and medical genetics: 25 years of my experience. *J Hum Genet.* 2009;54:1-8.
11. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA.* 2008;299:1335-1344.
12. Vink JM, Boomsma DI. Gene finding strategies. *Biol Psychol.* 2002; 61:53-71.
13. Rodríguez-Elías AK, Maldonado-Murillo K, López-Mendoza LF, Ramírez Bello J. Genetics and genomics in rheumatoid arthritis (RA): an update. *Gac Med Mex.* 2016;152:218-227.
14. Teruel M, Alarcón-Riquelme ME. The genetic basis of systemic lupus erythematosus: What are the risk factors and what have we learned. *J Autoimmun.* 2016;74:161-175.
15. Terali K, Ergören MÇ. The contribution of NOS3 variants to coronary artery disease: a combined genetic epidemiology and computational biochemistry perspective. *Int J Biol Macromol.* 2018;123:494-499.
16. Carreira PE, González-Crespo MR, Ciruelo E, Pablos JL, Santiago B, Gómez-Camara A, et al. Polymorphism of the interleukin-1 receptor antagonist gene: a factor in susceptibility to rheumatoid arthritis in a Spanish population. *Arthritis Rheum* 2005; 52: 3015-3019.

17. AlFadhli S, Ghanem AA. Influence of human DN1 VNTR polymorphism on DNASE1 expression in systemic lupus erythematosus and rheumatoid arthritis. *Immunol Invest.* 2014;43:411-423.
18. Yigit S, Inanir A, Tekcan A, Tural E, Ozturk GT, Kismali G, et al. Significant association of interleukin-4 gene intron 3 VNTR polymorphism with susceptibility to knee osteoarthritis. *Gene.* 2014;537:6-9.
19. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol.* 2013;9:141-153.
20. Cornélis F, Fauré S, Martinez M, Prud'homme JF, Fritz P, Dib C, et al. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc Natl Acad Sci U S A.* 1998;95:10746-10750.
21. Willems T, Gymrek M, Highnam G, Mittelman D, Erlich Y, 1000 Genomes Project Consortium. The landscape of human STR variation. *Genome Res.* 2014;24:1894-1904.
22. Aronin N, Kim M, Laforet G, DiFiglia M. Are there multiple pathways in the pathogenesis of Huntington's disease? *Philos Trans R Soc Lond B Biol Sci.* 1999;354:995-1003.
23. Roos CA, Tabrizi SJ. Huntington's disease: from molecular pathogenesis to clinical treatment. *Lancet Neurol.* 2011;10:83-98.
24. Paulson H. Machado-Joseph disease/spinocerebellar ataxia type 3. *Handb Clin Neurol.* 2012;103:437-449.
25. Rohilla KJ, Gagnon KT. RNA biology of disease-associated microsatellite repeat expansions. *Acta Neuropathol Commun.* 2017;5:63.
26. Wang J, Huang M, Zhao H, Wang C, Zhao X, Jiang X, et al. A novel tetranucleotide repeat polymorphism within KCNQ1OT1 confers risk for hepatocellular carcinoma. *DNA Cell Biol.* 2013;32:628-634.
27. Zhao X, Hu S, Wang L, Zhang Q, Zhu X, Zhao H, et al. Functional short tandem repeat polymorphism of PTPN11 and susceptibility to hepatocellular carcinoma in Chinese populations. *PLoS One.* 2014;9:e106841.
28. Yue CM, Bi MX, Tan W, Deng DJ, Zhang XY, Guo LP, et al. Short tandem repeat polymorphism in a novel esophageal cancer-related gene (ECRG2) implicates susceptibility to esophageal cancer in Chinese population. *Int J Cancer.* 2004;108:232-236.
29. Hsu LM, Huang YS, Yang SY, Chang FY, Lee SD. Polymorphism of T-cell receptor gamma short tandem repeats as a susceptibility risk factor of hepatocellular carcinoma. *Anticancer Res.* 2006;26:3787-3791.
30. Dziedzieiko V, Kurzawski M, Safranow K, Ossowski A, Piatek J, Parafiniuk M, et al. CAG repeat polymorphism in the androgen receptor gene in women with rheumatoid arthritis. *J Rheumatol* 2012; 39: 10-17.
31. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature.* 2009;461:747-753.
32. Mayeux R. Mapping the new frontier: complex genetic disorders. *J Clin Invest.* 2005;115:1404-1407.
33. Hanchard NA. Genetic susceptibility and single-nucleotide polymorphisms. *Semin Fetal Neonatal Med.* 2005;10:283-289.
34. Lewis CM, Knigh J. Introduction to genetic association studies. *Cold Spring Harb Protoc.* 2012;2012:297-306.
35. Alemán-Ávila I, Jiménez-Morales M, Beltrán-Ramírez O, Barbosa-Cobos RE, Jiménez-Morales S, Sánchez-Muñoz F, et al. Functional polymorphisms in pre-miR-146a and pre-miR499 are associated with systemic lupus erythematosus but not with rheumatoid arthritis or Graves' disease in Mexican patients. *Oncotarget.* 2017;8:91876-91886.
36. Rincón JF, Cano DL, Morales SJ, Jiménez ML, Cobos RE, Bello JR. The functional PTPN22 C1858T polymorphism confers risk for rheumatoid arthritis in patients from Central Mexico. *Clin Rheumatol.* 2016;35:1457-1462.
37. López Cano DJ, Cadena-Sandoval D, Beltrán-Ramírez O, Barbosa-Cobos RE, Sánchez-Muñoz F, Amezcua-Guerra LM, et al. The PTPN22 R263Q polymorphism confers protection against systemic lupus erythematosus and rheumatoid arthritis, while PTPN22 R620W confers susceptibility to Graves' disease in a Mexican population. *Inflamm Res.* 2017;66:775-781.
38. Mao L, Fang Y, Campbell M, Southerland WM. Population differentiation in allele frequencies of obesity-associated SNPs. *BMC Genomics.* 2017;18:861.
39. Wheeler E, Huang N, Bochukova EG, Keogh JM, Lindsay S, Garg S, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet.* 2013;45:513-517.
40. Chen L, Zhou W, Zhang L, Zhang F. Genome architecture and its role in human copy number variation. *Genomics Inform.* 2014;12:136-144.
41. Riggs ER, Ledbetter DH, Martin CL. Genomic variation: Lessons learned from whole-genome CNV analysis. *Curr Genet Med Rep.* 2014;2:146-150.
42. Choy KW, Setlur SR, Lee C, Lau TK. The impact of human copy number variation on a new era of genetic testing. *BJOG.* 2010;117:391-398.
43. Martin CL, Kirkpatrick BE, Ledbetter DH. Copy number variants, aneuploidies, and human disease. *Clin Perinatol.* 2015;42:227-242.
44. Mullaney JM, Mills RE, Pittard WS, Devine SE. Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet.* 2010;19:R131-R136.
45. Bhangale TR, Stephens M, Nickerson DA. Automating resequencing-based detection of insertion-deletion polymorphisms. *Nat Genet.* 2006;38:1457-1462.
46. Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, et al. Natural genetic variation caused by small insertions and deletions in the human genome. *Genome Res.* 2011;21:830-839.
47. Steri M, Orrù V, Idda ML, Pala M, Sidore C, Faà V, et al. Overexpression of the cytokine BAFF and autoimmunity risk. *N Engl J Med.* 2017; 376:1615-1626.
48. Say YH. The association of insertions/deletions (INDELs) and variable number tandem repeats (VNTRs) with obesity and its related traits and complications. *J Physiol Anthropol.* 2017;36:25.
49. Marques D, Ferreira-Costa LR, Ferreira-Costa LL, Correa RDS, Borges AMP, Ito FR, et al. Association of insertion-deletions polymorphisms with colorectal cancer risk and clinical features. *World J Gastroenterol.* 2017;23:6854-6867.
50. Maddhuri S, Bandaru S, Bhukya C, Cingeetham V, Malempati AR, Deepika MLN, et al. Association of CnB 5I/5D promoter gene polymorphism and serum calcineurin levels in early onset of coronary artery disease of south Indian cohort. *Gene.* 2017;632:1-6.
51. Choi J, Xu M, Makowski MM, Zhang T, Law MH, Kovacs MA, et al. A common intronic variant of PARP1 confers melanoma risk and mediates melanocyte growth via regulation of MITF. *Nat Genet.* 2017;49:1326-1335.
52. Wang C, Li L, Yin Z, Zhang Q, Zhao H, Tao R, et al. An indel polymorphism within pre-miR3131 confers risk for hepatocellular carcinoma. *Carcinogenesis.* 2017;38:168-176.
53. Jiang ZS, Hao ZH. An insertion/deletion polymorphism within the 3'-untranslated region of COL1A2 confers susceptibility to osteoporosis. *Mol Med Rep.* 2016;14:4415-4421.
54. Hachiya Y, Kawasaki A, Oka S, Kondo Y, Ito S, Matsumoto I, et al. Association of HLA-G 3' untranslated region polymorphisms with systemic lupus erythematosus in a Japanese population: A case-control association study. *PLoS One.* 2016;11:e0158065.